DOCUMENT RESUME

ED 148 262 HE 009 532

AUTHOR McShane, Michael G.; Sherman, Charles R.

TITLE Exploratory Analysis of Medical School Characteristics and Empirical Classifications of Medical Schools. Two Replications. Final Report.

INSTITUTION Association of American Medical Colleges, Washington, D. C.

SPONS AGENCY Health Resources Administration (DHEW/PHS), Bethesda, Md. Bureau of Health Manpower.
PUB DATE Dec 77
CONTRACT HRA-231-76-0011
NOTE 54p.; Not available in hard copy due to marginal legibility of original
AVAILABLE FROM Association of American Medical Colleges, One Dupont Circle, Washington, D.C. 20036

EDRS PRICE MF-$0.83 Plus Postage. HC Not Available from EDRS.
DESCRIPTORS Cluster Analysis; *Comparative Analysis; Curriculum; Educational Finance; Facilities; Factor Analysis; Faculty; Higher Education; *Medical Education; *Medical Schools; *Medical Students; *Professional Education; *Statistical Studies; Student Characteristics; Teacher Characteristics
IDENTIFIERS *College Characteristics

ABSTRACT

The report is one of four that describes the 1977 phase of an ongoing series of studies by the association. The purpose of the studies is to examine the interrelationships among variables and institutions based on the descriptive measures contained in the Institutional Profile System, which contains a vast array of data on medical schools describing students, faculty, curricula, facilities, finances, and other institutional characteristics. The two studies described in this report are replications of studies performed in 1976. The results of the replications were similar but not identical to earlier results. The results and the statistical measures used are discussed. Data are presented in tabular form. (MSE)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED148262

EXPLORATORY ANALYSIS OF MEDICAL SCHOOL CHARACTERISTICS AND EMPIRICAL CLASSIFICATIONS OF MEDICAL SCHOOLS

TWO REPLICATIONS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

FINAL REPORT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Assoc of American Medical Colleges

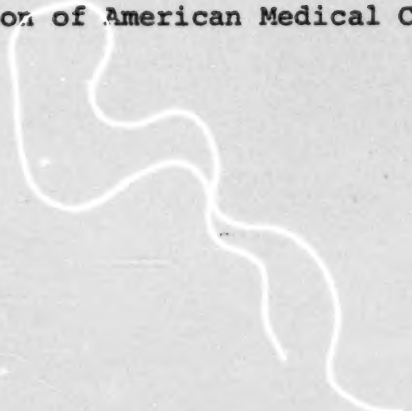
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM"

Association of American Medical Colleges
One Dupont Circle, N.W., Washington, D.C. 20036

U.S. Department of Health, Education, and Welfare
Public Health Service
Health Resources Administration
Bureau of Health Manpower
Contract No. 231-76-0011

HE 004532

© Association of American Medical Colleges, 1977



The Government retains the right to use, duplicate or disclose the contents of this report and to have or permit others to do so.

EXPLORATORY ANALYSIS OF MEDICAL SCHOOL CHARACTERISTICS AND EMPIRICAL CLASSIFICATIONS OF MEDICAL SCHOOLS

TWO REPLICATIONS

Michael G. McShane, Ph.D.
Charles R. Sherman, Ph.D.

DIVISION OF OPERATIONAL STUDIES

ASSOCIATION OF AMERICAN MEDICAL COLLEGES

FINAL REPORT

December 1977

The work upon which this publication is based was supported in part by the Bureau of Health Manpower, Department of Health, Education, and Welfare, pursuant to contract number 231-76-0011. However, any conclusions and/or recommendations expressed herein do not necessarily represent the views of the supporting agency.

CONTENTS

Page

List of Tables ii

List of Figures iii

Executive Summary v

Chapter I. Introduction 1
Background
Present Study

Chapter II. Method 3
Data Selection and Preparation
Factor Analysis
Cluster Analysis

Chapter III. Results 9
1976 Factor Pattern
1977 Factor Pattern in 1976 Variable Order.
1977 Factor Pattern with Re-ordered Rows
Cluster Analysis Results
Comparison with 1976 Results

Chapter IV. Summary and Conclusions. 31

Bibliography 33

Appendices 35
A. Abbreviations Used in Variable Labels . 35
B. Interpretation of the Factor Pattern Matrix 37
C. Results of Hierarchical Cluster Analysis Analysis 39

LIST OF TABLES

		Page
Table 1	Means, Standard Derivations, Numbers of Cases and Units for 33 Selected Medical School Variables, 1976 and 1977.	5
Table 2	1976 Factor Pattern Matrix Resulting from Analysis of 33 Variables Describing Medical Schools by Method of Principal Components with Varimax Rotation.	10
Table 3	1977 Factor Pattern Matrix from Analysis of Updated 1976-Selected Variables by Method of Principal Components and Varimax Rotation; Rows in 1976 Variable Order	12
Table 4	1977 Factor Pattern Matrix from Analysis of Updated 1976-Selected Variables by Method of Principal Components and Varimax Rotation; Rows Sorted by Empirical Groupings . . .	15
Table 5	Mean Values for 8 Clusters on Selected Variables Contributing to Factor Scores Used as the Basis for Cluster Analysis .	24

LIST OF FIGURES

		Page
Figure 1	Cluster Membership and Cluster Profiles on Six Factor Scores	20
Figure 2	Comparison of Cluster Membership Between 1976 Study and 1977 Replication	28

EXECUTIVE SUMMARY

This report, Exploratory Analysis of Medical School Characteristics and Empirical Classifications of Medical Schools - Two Replications, is one of four reports which describe the 1977 phase of an ongoing series of studies performed by the Association of American Medical Colleges (AAMC). The purpose of this series of studies is to examine the interrelationships among variables and institutions based on the descriptive measures contained in the Institutional Profile System (IPS), which contains a vast array of data on medical schools describing students, faculty, curricula, facilities, finances, and other institutional characteristics.

The two studies described in this report are replications of two studies performed in 1976.* Essentially, the same methods that were used in 1976 were repeated in 1977 on the same set of variables used in 1976, but in most cases, with more recent data. The data used in 1976 were primarily from 1974-75; those used in these studies primarily from 1975-76. The results of the present study showed nine primary dimensions (factors) of institutional differences. New variables based on these nine dimensions were computed for 111 schools, and these schools were formed into eight groups based on their similarities on six of the nine new variables using multivariate cluster analysis.

The results of the 1977 replications were similar but not identical to the 1976 results. Possible reasons for the lack of perfect correspondence among variables in factor structures and schools in clusters from one year to the next are discussed. They seemed to be caused by the initial selection of the variables, the stringency of the replication test, and actual changes in the relationships among the variables and among the schools.

---

*A Second Exploratory Analysis of the Relations Among Institutional Variables, (Sherman, 1977).

An Empirical Classification of U.S. Medical Schools by Institutional Dimensions, (McShane, 1977).

Chapter I

INTRODUCTION

Background

Since 1975, the Association of American Medical Colleges (AAMC) has conducted an ongoing series of studies examining quantified characteristics of medical schools in the United States. Available data from a number of different sources are routinely collected and stored by the AAMC. The major portion of the data used in these studies is accumulated annually through the student application and institutional research activities of the AAMC.

Using multivariate statistical methods, the studies in this series examine and reexamine several questions: In what basic ways are medical schools similar and different? What groups of schools are similar to one another? What is the global picture of institutional similarity with respect to specific characteristics?

In 1976, the AAMC conducted and reported A Second Exploratory Analysis of the Relations Among Institutional Variables, (Sherman, 1977). Thirty-three variables were selected, through a series of preliminary data analyses, to represent the spectrum of ways in which available data showed the faculty, students, curriculum and finances of medical schools to vary. A principal component analysis of the intercorrelations of these variables led to the identification of eight basic factors or dimensions of variation that subsumed most of the thirty-three measured characteristics. These were described as (1) graduate medical program emphasis, (2) size and age of the undergraduate medical program, (3) type of control (public versus private), (4) stage of institutional development, (5) research emphasis, (6) success in competitive research funding, (7) enrollment of minorities, and (8) the use of electives in the curriculum.

Scores on a continuous scale were derived for each school on each of the first six basic factors named above. Schools were clustered into eight groups based on the overall similarities of their profiles on the

six factor scores. The result was An Empirical Classification of U.S. Medical Schools by Institutional Dimensions, (McShane, 1977).

Present Study

The present study is an attempt to replicate the findings of the two previous studies, cited above, using the same methods and measures but newer data. The resulting factor pattern and institutional groupings are compared with the earlier results. The purpose of these comparisons is to assess the stability and, thereby, the reliability of the results. A stable outcome would provide assurance that the methods and data could reliably be used to assess relative changes in the functioning of individual medical schools.

II. METHOD

The scope of the analyses described in the report covered two phases of the analyses conducted in 1976. As much as possible, the same variables used in the earlier studies were extracted from AAMC's Institutional Profile System (IPS) with data one year more recent, and the same analyses performed in 1976 were repeated on the new data. There were three steps to the analysis: (1) selection of data elements, (2) factor analysis, and (3) cluster analysis. Each of these steps will be described in this chapter.

Data Selection and Preparation

During 1977, the size of IPS increased from approximately 8,000 to over 11,500 variables. Much of the data added to IPS were repetitions of data gathered in earlier years on annual questionnaires such as the Liaison Committee on Medical Education questionnaire (Parts I and II), the Fall Enrollment Questionnaire, and the Curriculum Directory questionnaire. However, many data were added from special surveys such as the Primary Care Questionnaire, and the Survey of Residencies conducted in 1977, and from other AAMC information systems. A large body of data on applicants and matriculants was aggregated by institution from the Medical Student Information System (MSIS), and extensive data on medical school faculties was obtained from the Faculty Roster Systems (FRS). Additional data were also obtained from the AMA's directory of Medical School Alumni (Martin, 1975), and from the Division of Research Grants (DRG) of NIH, although the former had not been updated since the previous studies. As these new sources of data were added to IPS, data elements were selected, extracted and compiled into a researchable data base for use in studies requiring institutional data. In addition to the "raw" measures extracted from IPS, the researchable data base contained a number of derived or computed variables, which measured relative characteristics of the institutions. A complete list of the variables in the 1977 Researchable Data Base may be found in a descriptive report by Agro, et al. (1977).

Among the 950 variables included in the researchable data base were the 33 variables used in the 1976 studies of the interrelationships among variables and institutions that are replicated in the report. A listing of the

variables, with mean values and the number of schools for which data were available, is presented in Table 1. (Appendix A is a glossary of abbreviations used in the labels). For the majority of the variables, the same variable used in the 1976 studies was derived with data one year more recent. Most of the variables used in the 1976 studies were based on data from the 1974-75 academic year, and in the current studies, on the 1975-76 academic year.

There were three exceptions to the general method of determining the variables to be used described above. One variable, the amount of change (expressed in percentage of constant dollars) in sponsored research monitored by DRG between 1967-69 and 1972-74 (INC003), used the same data as in the 1976 studies. There was no more recent data available for extending this particular measure an additional year. Also, since the AMA data on alumni were not revised between 1976 and 1977, no new data were available for computing the percentage of alumni in general practice (STC105) or the percentage of alumni who were board certified (STC112). Rather than use the same data for these two variables, however, the percentages were computed for the subset of alumni who graduated between 1960 and 1969 rather than all alumni. The intent of the change was to avoid repetition and, secondarily, produce a better measure of the career patterns of recent alumni of the medical schools. The result of this change, however, was an equivalent mean percentage of alumni board certified but less variation among the schools. Also, the mean percentage of alumni in general practice dropped from 14.3 to 10.9 percent.

The mean values for the other 30 variables remained fairly consistent from one year to the next, but several changes of interest were noted. The mean number of medical students enrolled, the mean adjusted total revenue, the mean percentage of students having undergraduate GPA's between 3.6 and 4.0, the mean percentage of female first-year medical students, and the mean MCAT Science scores for first year medical students all rose slightly in one year.

Factor Analysis

The second step in the analytic process employed in the study was to factor analyze the 33 variables described above. The purpose of the factor analysis was to examine the degree of stability in the interrela-

TABLE 1

MEANS, STANDARD DEVIATIONS, NUMBERS OF CASES AND UNITS FOR 33 SELECTED MEDICAL SCHOOL VARIABLES, 1976 AND 1977

VARIABLE	LABEL	MEAN 1976	MEAN 1977	STANDARD DEVIATION 1976	STANDARD DEVIATION 1977	NUMBER OF CASES 1976	NUMBER OF CASES 1977	UNITS
FAR053	FSS: AV SAL SFT ASSOC PROF BAS SCI	24.84	26.12	2.87	3.19	104	104	THOUSANDS OF DOLLARS
STC043	RAT: HOUSESTAFF TO UNDERGRAD MD STUD	.83	.81	.72	.56	114	114	HOUSESTAFF PER STUDENT
INC058	RAT: MD STUDENTS TO FT FAC	1.76	1.72	.84	.81	113	114	STUDENT PER FACULTY
STC105	AMA: % 60-69 ALUMNI IN GENERAL PRACTICE[1]	14.29	10.86	7.18	7.84	98	82	%
PAC001	FRS - % FT SAL FAC WITH MD OR MD - PHD[2]	62.92	61.55	12.12	12.40	114	112	%
STR009	# UNDERGRADUATE MEDICAL STUDENTS	474.35	493.37	224.00	222.50	114	114	STUDENTS
INC048	LOG AGE OF MEDICAL SCHOOL	1.64	1.67	.55	.52	117	118	LOG (CURRENT YEAR MINUS YEAR FOUNDED)
STC112	AMA: % 60-69 ALUMNI BOARD CERTIFIED[3]	45.97	45.86	19.98	8.83	98	84	%
INR002	CONTROL: 0 = PUBLIC, 1 = PRIVATE	.40	.40	.49	.49	117	119	NOMINAL (0,1)
STR432	1976-77 RESIDENT TUITION	2178.54	2493.75	1485.05	1917.86	112	102	DOLLARS
STC029	% IN-STATE 1ST YR MD STUD	74.58	75.35	27.13	27.56	112	109	%
STC084	RAT: APPLICANTS PER 1ST-YR MD STUD	25.94	26.78	16.35	18.72	114	116	APPLICANTS PER STUDENT
INC007	% REV FROM FED SOURCES + RCOV IND COSTS	37.33	34.72	13.17	12.03	106	104	%
INC012	% REV FROM GIFTS + NON-GOVT SPONS PGMS[4]	6.74	6.69	5.63	5.03	105	107	%
STC082	% UNDERREP MINORITY 1ST YR MD STUD	9.50	10.99	10.07	15.34	102	114	%
FAC004	FRS: % FT SAL FAC FR ETHNIC MINORITIES[5]	3.26	3.21	10.57	10.58	114	114	%
STC008	% NON-US-CANADIAN 1ST-YR MD STUD	1.29	1.20	2.05	1.99	113	111	%
STR371	MEAN MCAT SCIENCE SCORES OF MATRC. TOTAL	601.31	617.01	34.75	32.56	114	116	POINTS
INR141	NIH RO1 GRANTS: $ AWARD AS % OF APP SBMT	26.30	23.35	14.62	13.58	114	112	%
INR142	NIH RO1 GRANTS: MEAN STD PRIORITY SCR	.08	.09	.33	.48	107	111	STANDARDIZED SCORE
INR137	NIH RO1 GRANTS: APPROVAL RATE OF APPLS	69.01	67.47	21.93	19.40	114	114	%
STC003	% FEMALE MD STUDENTS	18.63	20.95	7.04	6.93	114	114	%
CRR028	REL ELECTIVES: ALCOHOLISM	.63	.70	.49	.46	113	115	NOMINAL (1 = AVAILABLE, 0 = NA)
CRC002	# OF MED-RELATED ELECTIVES OFFERED	7.58	7.93	4.03	3.89	117	115	ELECTIVE COURSES
FAC019	RAT: VOL FAC TO FT FAC	2.35	2.35	2.09	1.83	111	109	FACULTY PER FACULTY
INC003	DRG FED SPON RES CON$ CHG 67-9 TO 72-4[6]	42.97	42.97	107.16	107.16	101	101	ANNUAL % CHANGE (CONST. $)
STC114	PROJTD ANNL % 1ST-YR ENROLL CHG 1975-80[7]	2.17	1.46	5.37	4.34	110	105	ANNUAL % CHANGE
INR136	NIH RO1 GRANTS: # APPLS APPROVED	27.91	24.02	21.51	18.19	114	115	APPLICATIONS
INC026	% EXPD FOR ADMIN & GENL EXPENSE	10.33	10.73	5.71	5.28	111	112	%
INC017	% TOTAL EXPD FOR SPON RESEARCH	21.78	21.02	12.22	11.99	111	112	%
STC045	RAT: BMS GRAD STUD TO UNDERGRAD MD STUD	.22	.27	.18	.20	114	114	STUDENT PER STUDENT
INC004	ADJUSTED TOTAL REVENUE THOUSAND $	27177.31	29982.29	17923.09	19537.82	111	112	THOUSANDS OF DOLLARS
STC013	% 1ST-YR MD STUD: PRE-MED GPA 3.6 - 4.0	37.61	43.36	18.71	16.90	113	109	%

[1]Variable for 1976 was % living MD-Alumni in general practice
[2]Variable for 1976 was % PT & FT SAL FAC with MD
[3]Variable for 1976 was % living MD Alum Board certified
[4]Variable for 1976 was mislabelled % REV from all gifts
[5]Variable for 1976 was % PT & FT SAL PAC from ethnic minorities
[6]Data for this variable were exactly the same in both years
[7]In 1976 this variable referred to the years 1974-79

tionships among the variables, and to note the differences in the patterns of relationships brought about by the changes noted above.

Principal component analysis was used to reduce the 33 variables to the primary dimensions being assessed. Factor analysis, the family of analytic techniques of which principal components is a member, is based on the presumption that the variables do not each measure a unique element of the institution known as a medical school, but rather that the variables combine to assess dimensions on which the schools can be measured. Ten components, accounting for 75.6 percent of the total variance in the data, were extracted. Three varimax rotations, involving 10, 9, and 8 factors respectively, were performed. The rotated factor patterns were examined for compatability to 1976 results and intrinsic meaning, and one solution was selected for presentation.

Cluster Analysis

The final stage of the analysis for this study was to use factor scores from the 9 dimensions derived in the factor analysis to cluster the medical schools. In much the same way that factor analysis is used to examine patterns of correlation among variables, cluster analysis is used to examine patterns of similarity among objects, in this case institutions, across a number of variables. Just as factor analysis derives groups of variables that are similar to one another and different from those in other groups, cluster analysis groups together objects (schools) that are similar to each other and differ in some way from schools in other groups.

As in the earlier studies, the cluster analysis was performed in two steps. The first step was a hierarchical cluster analysis using the method developed by Ward (1963). Hierarchical cluster analysis techniques initially view each object as belonging to a cluster of size 1. In the first step of the analysis the two most similar groups are combined to form a single, larger group of size 2. In each subsequent step, the two most similar groups are combined into a single group. One limitation of hierarchical cluster analysis is that once a combining of two groups has taken place, the

objects which make up that group remain joined for the duration of the analysis. By forcing all objects to be combined and remain intact once they are combined, hierarchical cluster analysis may cause distortions of natural clusters by the inclusion of outlying objects.

To overcome the artifacts of hierarchical cluster analysis, the hierarchical solution was refined through the use of a second step, non-hierarchical cluster analysis, which places objects into a predetermined number of clusters in such a way that a specified criterion is optimized. Non-hierarchical cluster analysis allows objects to move among clusters until a "best" solution is found, and thus reduces the distorting effect of outlying objects on cluster membership. Non-hierarchical cluster analysis, however, does require specification of initial estimates of the location of the centers of the clusters as well.

In this study, a non-hierarchical cluster analyses technique developed by Forgy (1965) was used. By this method, using the number of clusters and cluster centroids specified by the user, each object is assigned to the cluster which has the closest centroid. After all objects have been initially assigned to clusters, new cluster centroids are computed for each cluster based on the objects assigned to the cluster. The distance from each object to the centroids of each cluster is then computed and objects are reassigned, if necessary, to the cluster which now has the closest centroid. After the reassignment of objects, the cluster centroids are recomputed and a new cycle of computing distances, assigning objects to clusters, and recomputing cluster centroids is begun. The cycle is repeated until no objects are reassigned after cluster centroids have been calculated. This procedure, like the Ward approach, minimizes the differences among the objects within a cluster, but allows for the reassignment of objects to clusters.

In the present study, factor scores were computed for the nine factors derived by the factor analysis. Several combinations of factor scores were used as input to hierarchical cluster analysis. By comparing the results of the hierarchical clustering solution, it was determined that the six factor score combination presented clear delineations among the schools which seemed to resemble those derived in the 1976 studies. Therefore, the six factor scores and seedpoints resulting from the hierarchical cluster analysis based on

those six factor scores were used as input to the non-hierarchical cluster analysis, and an eight-cluster solution was obtained. The results of this analysis, along with the results of the factor analysis and a comparison between the results obtained in 1977 and those obtained using the same variables one year earlier, are presented in the next chapter.

Chapter III

RESULTS AND DISCUSSION

The analyses replicated in this study have two major stages. The first stage is a principal components analysis leading to identifiable descriptive factors. The second stage is a cluster analysis of schools, based on the descriptive factors resulting from the first stage. Each major stage consists of a number of steps. The results of the replications of both principal components analysis and cluster analysis are presented in this chapter.

The result of a principal components analysis is a set of numbers called "factor loadings." The absolute values of the loadings represent the degree to which individual variables belong to each of several groupings of variables. The set of factor loadings are arrayed in a "factor pattern matrix" having one row for each variable and one column for each "principal component" or grouping of variables interpreted to describe a basic "factor." Three factor pattern matrices are presented in this chapter: first, the matrix resulting from the 1976 study; second, the matrix resulting from the 1977 replication with the rows in the order corresponding to the 1976 ordering of the variables; third, the 1977 results with the rows sorted to enhance interpretation. (A brief set of guidelines by which to interpret a factor pattern matrix is given in Appendix B).

The result of a cluster analysis is a list of groups of medical schools, where each school belongs only to the one group of schools to which it is "most" similar. The group memberships derived in this study will be presented and compared with the group memberships found in the 1976 study.

1976 Factor Pattern

Table 2 presents the 1976 factor pattern matrix. The large numbers (in absolute value) have been accentuated by "boxes," moderate values by asterisks. The 33 rows are labeled with the 1976 variable names. The first eight columns of numbers are labeled according to the factor described by variables having large

TABLE 2

1976 FACTOR PATTERN MATRIX RESULTING FROM ANALYSIS OF 33 VARIABLES DESCRIBING U.S. MEDICAL SCHOOLS BY METHOD OF PRINCIPAL COMPONENTS WITH VARIMAX ROTATION

	VARIABLE	1 Graduate Medical Program	2 Size, Age	3 Control	4 Minority	5 Research Funding Success	6 Curriculum Electives	7 Development Stage	8 Research Emphasis	h^2
1	VAR388 AV SALARY - SFT ASSOC PROF BASIC SCIENCE	.84	-.02	.03	-.00	-.01	.05	.16	-.03	.73
2	STC043 RAT: HOUSESTAFF TO UNDERGRAD MD-STUD	.79	-.03	.19	-.03	-.00	.04	.05	.07	.68
3	INC058 RAT: MD STUDENTS TO FT FAC	-.67	.22	-.02	-.05	-.14	.05	.23	-.36*	.71
4	STC105 % LIVING MD-ALUMNI IN GENERAL PRACTICE	-.54	.33	-.13	.04	-.42*	-.24	.14	-.14	.70
5	FAC001 % PT & FT SAL FAC WITH MD	.40	.27	.33*	.06	.14	-.02	-.01	-.03	.37
6	VAR016 # MD-STUDENTS	-.09	.88	-.06	.03	-.04	.08	-.04	.16	.83
7	INC048 LOG AGE OF MEDICAL SCHOOL	-.28	.75	.21	.01	.15	.03	-.33*	.15	.83
8	STC112 % LIVING MD ALUM BOARD CERTIFIED	.14	.71	.32	-.17	.04	.07	-.28	.09	.75
9	VAR002 CONTROL: 0 = PUBLIC, 1 = PRIVATE	.15	.14	.87	.09	-.00	.03	-.13	-.01	.83
10	VAR394 1975-76 RESIDENT MD-STUDENT TUITION	.05	.13	.86	-.07	.13	.10	-.14	-.02	.82
11	STC029 % IN-STATE 1ST-YR MD-STUD	.01	-.06	-.81	-.23	-.14	.00	.18	-.16	.79
12	STC084 RAT: APPLICANTS PER 1ST-YR MD-STUD	.10	-.09	.79	-.03	.01	-.07	.27	-.03	.72
13	INC007 % REV FROM FED SOURCES & RCOV INC COSTS	.05	-.01	.48	.05	.22	.28	-.27	.48*	.66
14	INC012 % REV FROM ALL GIFTS	.20	.08	.38	.11	-.32	.10	-.06	.13	.33
15	STC082 % UNDERREP MINORITY 1ST-YR MD-STUD	-.04	-.09	.06	.94	-.10	.02	.06	-.03	.91
16	FAC004 % PT & FT SAL FAC FROM ETHNIC MINORITIES	-.11	-.06	-.04	.87	.03	.03	-.14	-.14	.82
17	STC008 % NON US-CANADIAN 1ST-YR MD-STUD	.19	.17	.25	.67	-.08	-.03	.06	.10	.60
18	VAR093 1ST-YR MD-STUD: MEAN MCAT SCIENCE SCORE	.43*	.23	.35*	-.44	.26	.04	.08	.36*	.75
19	INC046 NIH-NIMH R01 $ AWARD AS % OF $ APP SBMT	-.01	.11	.14	-.10	.84	.04	.07	.13	.77
20	VAR352 IMPAC: MEAN STD P-SCR - R01 APP	-.35	.04	-.09	.14	-.73	-.05	.21	-.05	.74
21	INC045 IRG APPROVAL RATE OF NIH R01 COMP APPS	-.04	.29	-.05	.01	.70	-.03	.22	.38*	.78
22	STC003 % FEMALE MD STUDENTS	.20	-.13	.18	.31	.48	.24	.02	-.28	.56
23	VAR273 REL ELECTIVES: ALCOHOLISM	.07	.03	-.01	.02	-.03	.88	.02	.03	.79
24	CRC002 # OF RELATED ELECTIVES OFFERED	.03	.14	.12	.01	.14	.82	.01	.24	.78
25	FAC019 RAT: VOL FAC TO FT FAC	-.12	-.02	-.02	-.11	.08	.10	.74	-.30	.68
26	INC003 DRG FED SPON RES CON$ %CHG 67-9 TO 72-4	.14	-.44*	-.12	.15	-.01	.00	.73	.17	.82
27	STC114 PROJTD ANNL % 1ST-YR ENROLL CHG: 1974-79	.09	-.43*	-.17	-.04	-.01	-.09	.58	-.17	.60
28	VAR384 DRG GRANTS - # R01 APPS APPROVED	.41*	.41*	.05	-.01	.27	.05	-.03	.67	.87
29	INC026 % EXPD FOR ADMIN & GENL EXPENSE	.19	-.13	.02	-.13	.15	-.02	.25	-.64	.57
30	INC017 % TOTAL EXPD FOR SPON RESEARCH	.24	.13	.45*	-.02	.20	.26	-.04	.63	.78
31	STC045 RAT: BMS GRAD-STUD TO UNDERGRAD MD-STUD	-.05	.01	.09	-.09	.19	.23	.01	.61	.48
32	INC004 ADJUSTED TOTAL REVENUE	.43*	.52	-.01	.04	.16	.05	-.08	.57	.82
33	STC013 % 1ST-YR MD-STUD: PRE-MED GPA 3.6-4.0	.23	.02	-.06	-.19	-.04	-.05	-.04	.55	.40
	COLUMN SUM OF SQUARES	3.36	3.21	4.01	2.63	2.71	1.85	2.13	3.36	
	PERCENT OF VARIANCE	14.44	13.81	17.23	11.30	11.65	7.95	9.15	14.46	

"loadings" in the corresponding column. The last column, labeled "h^2," contains the communality of each row. These values reflect the degree to which the information carried by each variable is contained in all rotated components.

In 1976, eight components, accounting for 70.5 percent of the total variance, had the most readily interpetable rotated pattern compared with the 9, 7, 6, 5 and 4 component solutions. Detailed interpretations of the factors are presented in the 1976 report.

1977 Factor Pattern in 1976 Variable Order

Table 3 presents the 1977 factor pattern matrix with rows corresponding to the rows, hence to the same variables, in Table 2. The variable labels describe the data used in the replication. (Differences between the 1976 and 1977 variables are discussed in Chapter II). Again, large and moderately large values have been accentuated by "boxes" and asterisks.

The nine component rotated pattern, accounting for 72.2 percent of the total variance, was selected from among the 10, 9 and 8 component solutions examined, as providing the most interpetable pattern and the best match to the 1976 matrix. The factor patterns are fairly similar. Several prominent differences, however, are immediately noticeable. Their causes are complex and are more likely to include changes in the variables used than changes in the medical education enterprise.

The new factor pattern has nine components compared with eight in 1976. Aside from other realignments of correlated variables, the research emphasis component has split into two sub-groups of variables. The ratio of basic science graduate students to undergraduate medical students and the percentage of total school expenditures for sponsored programs, principally research, no longer appear to be strongly related to four other variables thought to be related to large research programs. It may be that the new pattern points to a differentiation between small (in enrollment) schools that have a research emphasis (the ninth component) and large schools that have a research emphasis as well as a larger number of ongoing research projects. The new eighth component also has

TABLE 3
1977 FACTOR PATTERN MATRIX FROM ANALYSIS OF UPDATED 1976-SELECTED VARIABLES BY METHOD OF PRINCIPAL COMPONENTS AND VARIMAX ROTATION; ROWS IN 1976 VARIABLE ORDER

VARIABLE		VARIABLE GROUP (COMPONENT)									
		1	2	3	4	5	6	7	8	9	h^2
FAR053	FSS: AV SAL SFT ASSOC PROF BAS SCI	.76	.39*	-.05	-.04	.13	.01	.22	.00	-.06	.79
STC043	RAT:HOUSESTAFF TO UNDERGRAD MD STUDENTS	.81	-.04	.13	-.08	.20	.04	.10	.08	.19	.77
INC058	RAT: MD STUDENTS TO FT FAC	-.53	-.06	-.04	.00	-.05	.01	-.35	-.40*	-.39*	.72
STC105	AMA:% 60-69 ALUMNI IN GENL PRACTICE	-.08	-.82	-.36	-.12	-.04	-.01	-.09	.01	-.14	.86
FAC001	FRS-%FT SAL FAC WITH MD OR MD-PHD	.43	.13	.37	-.12	.26	.22	-.01	.17	-.33	.61
STR009	# UNDERGRADUATE MEDICAL STUDENTS	-.07	.08	-.08	-.17	-.05	.31	.54	.48*	-.31	.76
INC048	LOG AGE OF MEDICAL SCHOOL	-.25	.25	.23	-.06	.03	.06	.60	.46*	-.11	.77
STC112	AMA:% 60-69 ALUMNI BOARD CERTIFIED	.15	.79	.34	-.26	.08	.06	.06	.09	.12	.86
INR002	CONTROL: 0=PUBLIC, 1=PRIVATE	.17	.23	.83	.03	-.04	.05	.06	.09	-.10	.79
STR432	1976-77 RESIDENT TUITION	.00	.12	.87	-.12	.01	.07	.10	-.02	-.12	.81
STC029	% IN-STATE 1ST-YR MD STUDENTS	.08	-.24	-.84	-.11	.01	-.05	.09	-.09	-.12	.81
STC084	RAT:APPLICANTS PER 1ST-YR MD STUDENT	.21	.32	.69	-.06	-.09	.05	.09	-.24	.04	.70
INC007	% REV FROM FED SOURCES & RCOV IND COSTS	.17	-.01	.57	-.02	.14	.29	-.21	.17	.41*	.70
INC012	% REV FROM GIFTS & NON-GOVT SPONS PGMS	.62	.01	.18	.04	-.21	.18	-.36	-.09	.10	.64
STC082	% UNDERREP MINORITY 1ST YR MD STUDENTS	.02	-.02	.01	.88	-.01	.04	.04	.01	-.02	.78
FAC004	FRS: % FT SAL FAC FR ETHNIC MINORITITES	-.09	-.03	.03	.90	-.13	.14	-.04	.09	-.07	.87
STC008	% NON-US-CANADIAN 1ST-YR MD STUD	.00	-.19	.52	.12	.21	.04	-.04	.10	.16	.40
STR371	MEAN MCAT.SCIENCE SCORES OF MATRICULANTS	.20	.38	.23	-.52	.34	.16	-.00	.20	.24	.74
INR141	NIH R01 GRANTS: $ AWARD AS % OF APP SBMT	.04	-.15	.14	-.12	.80	-.02	.02	.07	.04	.70
INR142	NIH R01 GRANTS: MEAN STD PRIORITY SCORE	-.15	-.10	.01	.12	-.71	-.18	-.21	-.09	.08	.64
INR137	NIH R01 GRANTS: APPROVAL RATE OF APPLS	.00	.20	-.10	.03	.73	.00	-.10	.10	.29	.69
STC003	% FEMALE MD STUDENTS	.16	.32	.16	.40*	.39	.12	.03	-.30	.01	.58
CRR028	REL ELECTIVES: ALCOHOLISM	.13	.01	.07	.06	.03	.82	.15	-.10	-.03	.74
CRC002	# OF MED-RELATED ELECTIVES OFFERED	.01	.06	.22	.12	.11	.82	-.05	.04	.23	.81
FAC019	RAT: VOL FACULTY TO FT FACULTY	-.01	.11	-.12	-.16	.16	.17	.17	-.65	-.08	.56
INC003	DRG FED SPON RES CONS CHG 67-9 TO 72-4	.01	-.13	-.06	-.01	.09	.17	.82	-.06	.05	.73
STC114	PROJTD ANNL % 1ST-YR ENROLL CHG 1975-80	.04	.15	-.15	-.01	-.01	.01	.80	-.18	-.25	.78
INR136	NIH R01 GRANTS: # APPLS APPROVED	.33	.22	.03	-.20	.26	.29	-.16	.57	.39*	.85
INC026	% EXPD FOR ADMIN & GENL EXPENSE	.04	.08	-.05	-.07	-.22	.12	.04	-.71	-.01	.58
INC017	% TOTAL EXPD FOR SPONSORED RESEARCH	.31	.18	.36	-.11	.32	.31	-.16	.24	.39	.71
STC045	RAT:BMS GRAD STUD TO UNDERGRAD MD STUD	.13	.19	.05	-.17	.11	.12	-.00	.16	.73	.68
INC004	ADJUSTED TOTAL REVENUE THOUSAND $	.20	.31	.01	-.17	.14	.25	-.12	.71	.23	.82
STC013	% 1ST-YR MD STUD:PRE-MED GPA 3.6-4.0	.19	.07	-.05	-.35	.17	.15	-.24	.46	.30	.57
	COLUMN SUM OF SQUARES	2.71	2.40	4.04	2.50	2.53	2.08	2.59	3.01	1.96	
	PERCENT OF VARIANCE	11.36	10.09	16.97	10.51	10.61	8.75	10.87	12.63	8.22	

secondary contributions (moderate loadings in rows six and seven) from variables describing size and age of the institution.

The change of variables that describe the specialization and non-specialization of medical school alumni appears to have caused two changes. First, the percentage of recent (1960-69) alumni in general practice is not related to the relative emphasis of the medical school on its graduate medical programs. In the 1976 study, the percentage of total alumni who are general practitioners was observed to be inversely related to GME emphasis. Second, the percentage of recent alumni who became certified by one or more specialty boards is not related to size and age of the undergraduate medical program as was the percentage of total alumni. With the new focus on a relatively short (10 year) interval, the two percentage measures (G.P.'s and specialists) are more strongly inversely related to one another. In the 1977 replication they form a new principal component of their own (Column 2).

Variables describing the size and age of the undergraduate medical program realigned with two components in the 1977 analysis. They now appear to be more strongly related to the stage of development of the medical school, defined by recent changes in federally-sponsored research and projected growth in enrollment (Column 7). They are also related, although somewhat less strongly, to the scope of the ongoing sponsored research enterprise at the school (Column 8), older and larger schools being involved in more research.

The percentage of the medical school's revenue coming from gifts and non-government sponsored programs (foundations, voluntary health agencies, business and industry, alumni, and endowment) no longer appears to be related to whether a school is publicly or privately controlled. The apparent relationship in the 1976 was weak and was not replicated. In the 1977 data, the receipt of such revenues appears to be related to a school's emphasis on graduate medical programs.

The presence of non-Canadian foreign students no longer appears to be related to the presence of minority persons. The 1977 pattern suggests that private schools may now be more likely to accept foreign students. The low communality (.40) for this variable,

however, suggests that such speculation may have little reliability.

The presence of more volunteer faculty per full-time faculty appears to be characteristic of schools with small scale research activities rather than of schools in early stages of development. Again, a low communality for this ratio discourages strong speculation.

1977 Factor Pattern with Re-ordered Rows

The selection of variables in the 1976 study was the result of a series of principal components analyses with varying sets of variables and varying numbers of rotated components. The resulting pattern judged most interpretable was selected for presentation, discussion, and to serve as the basis for institutional classification.

Table 4 presents the replicated factor pattern with rows (variables) re-ordered to enhance interpretation of the 1977 matrix without reference to the 1976 results. Since the variables used in 1977 were selected solely because they were used in 1976 and not because they led to a most easily interpetable solution, it should not be surprising that the naming and interpretation of factors makes use of greater imaginative license. Greater freedom in the selection and substitution of variables would probably have led to more interpretable components. The overall pattern, however, is similar to the 1976 pattern. Prominent differences are discussed in the above section.

There are nine groups of variables that describe nine principal components of variation among medical schools. These factors are:

Factor 1. The first five variables have their most prominent loadings in column one. These variables are related, either directly or indirectly, to an institution's emphasis on a graduate medical program. There tend to be more housestaff (interns and residents per medical student) faculty salaries are higher. (While basic science and clinical faculty salaries are highly correlated across schools, basic science salaries were judged to be more reliable measures and,thus were used in these studies). There are fewer medical

TABLE 4

1977 FACTOR PATTERN MATRIX FROM ANALYSIS OF UPDATED 1976-SELECTED VARIABLES BY METHOD OF PRINCIPAL COMPONENTS AND VARIMAX ROTATION; ROWS SORTED BY EMPIRICAL GROUPINGS

VARIABLE		1 Graduate Medical Program	2 Alumni Specialization	3 Control	4 Minority Representation	5 Research Funding Success	6 Curriculum Electives	7 Development	8 Research Enterprise	9 Research Training	h^2
STC043	RAT:HOUSESTAFF TO UNDERGRAD MD STUDENTS	.81	-.04	.13	-.08	.20	.04	.10	.08	.19	.77
FAR053	FSS: AV SAL SFT ASSOC PROF BAS SCI	.76	.39*	-.05	-.04	.13	.01	.22	.00	-.06	.79
INC012	% REV FROM GIFTS & NON-GOVT SPONS PGMS	.62	.01	.18	.04	-.21	.18	-.36	-.09	.10	.64
INC058	RAT: MD STUDENTS TO FT FAC	-.53	-.06	-.04	.00	-.05	.01	-.35	-.40*	-.39*	.72
FAC001	FRS-%FT SAL FAC WITH MD OR MD-PHD	.43	.13	.37	-.12	.26	.22	.01	.17	-.33	.61
STC105	AMA:% 60-69 ALUMNI IN GENL PRACTICE	-.08	-.82	-.36	-.12	-.04	-.01	.09	.01	-.14	.86
STC112	AMA:% 60-69 ALUMNI BOARD CERTIFIED	.15	.79	.34	-.26	.08	.06	-.06	.09	.12	.86
STR432	1976-77 RESIDENT TUITION	.00	.12	.87	-.12	.01	.07	-.10	-.02	-.12	.81
STC029	% IN-STATE 1ST-YR MD STUDENTS	.08	-.24	-.84	-.11	.01	-.05	.09	-.09	-.12	.81
INR002	CONTROL: 0=PUBLIC, 1=PRIVATE	.17	.23	.83	.03	-.04	.05	-.06	.09	-.10	.79
STC084	RAT:APPLICANTS PER 1ST-YR MD STUDENT	.21	.32	.69	-.06	-.09	.05	.09	-.24	.04	.70
INC007	% REV FROM FED SOURCES & RCOV IND COSTS	.17	-.01	.57	-.02	.14	.29	-.21	.17	.41*	.70
STC008	% NON-US-CANADIAN 1ST-YR MD STUD	.00	-.19	.52	.12	.21	.04	-.04	.10	.16	.40
FAC004	FRS: % FT SAL FAC FR ETHNIC MINORITITES	-.09	-.03	.03	.90	-.13	.14	-.04	.09	-.07	.87
STC082	% UNDERREP MINORITY 1ST YR MD STUDENTS	.02	-.02	.01	.88	-.01	.04	.04	.01	-.02	.78
STR371	MEAN MCAT.SCIENCE SCORES OF MATRICULANTS	.20	.38	.23	-.52	.34	.16	-.00	.20	.24	.74
STC003	% FEMALE MD STUDENTS	.16	.32	.16	.40	.39*	.12	.03	-.30	.01	.58
INR141	NIH R01 GRANTS: $ AWARD AS % OF APP SBMT	.04	-.15	.14	-.12	.80	-.02	.02	.07	.04	.70
INR137	NIH R01 GRANTS: APPROVAL RATE OF APPLS	.00	.20	-.10	.03	.73	.00	-.10	.10	.29	.69
INR142	NIH R01 GRANTS: MEAN STD PRIORITY SCORE	-.15	-.10	.01	.12	-.71	-.18	-.21	-.09	.08	.64
CRR028	REL ELECTIVES: ALCOHOLISM	.13	.01	.07	.06	.03	.82	.15	-.10	-.03	.74
CRC002	# OF MED-RELATED ELECTIVES OFFERED	.01	.06	.22	.12	.11	.82	-.05	.04	.23	.81
INC003	DRG FED SPON RES CON$ CHG 67-9 TO 72-4	.01	-.13	-.06	-.01	.09	.17	.82	-.06	.05	.73
STC114	PROJTD ANNL % 1ST-YR ENROLL CHG 1975-80	.04	.15	-.15	-.01	-.01	.01	.80	-.18	-.25	.78
INC048	LOG AGE OF MEDICAL SCHOOL	-.25	.25	.23	-.06	.03	.06	-.60	.46*	-.11	.77
STR009	# UNDERGRADUATE MEDICAL STUDENTS	-.07	.08	-.08	-.17	-.05	.31	-.54	.48*	-.31	.76
INC004	ADJUSTED TOTAL REVENUE THOUSAND $	.20	.31	.01	-.17	.14	.25	-.12	.71	.23	.82
INC026	% EXPD FOR ADMIN & GENL EXPENSE	.04	.08	-.05	-.07	-.22	.12	.04	-.71	-.01	.58
FAC019	RAT: VOL FACULTY TO FT FACULTY	-.01	.11	-.12	-.16	.16	.17	.17	-.65	-.08	.56
INR136	NIH R01 GRANTS: # APPLS APPROVED	.33	.22	.03	-.20	.26	.29	-.16	.57	.39*	.85
STC013	% 1ST-YR MD STUD:PRE-MED GPA 3.6-4.0	.19	.07	-.05	-.35	.17	.15	-.24	.46	.30	.57
STC045	RAT:BMS GRAD STUD TO UNDERGRAD MD STUD	.13	.19	.05	-.17	.11	.12	-.00	.16	.73	.68
INC017	% TOTAL EXPD FOR SPONSORED RESEARCH	.31	.18	.36	-.11	.32	.31	-.16	.24	.39	.71
	COLUMN SUM OF SQUARES	2.71	2.40	4.04	2.50	2.53	2.08	2.59	3.01	1.96	
	PERCENT OF VARIANCE	11.36	10.09	16.97	10.51	10.61	8.75	10.87	12.63	8.22	

students per faculty. Proportionately more faculty hold M.D. degrees. These schools appear to receive greater proportions of their revenue from gifts, foundations and other non-government sponsors.

Factor 2. Focusing on graduates of the decade from 1960 to 1969, the tendency of a school's graduates to seek specialized training appears to be independent of other derived factors. Secondary loadings, however, suggest that graduates of public medical schools are less likely to specialize. This factor has been labeled "alumni specialization."

Factor 3. Six variables are related to the public or private control of a medical school. Positive loadings reflect "privateness." Private schools tend to have higher tuitions, proportionately fewer in-state students, more applicants for each student opening, greater dependency on federal monies, and more foreign students.

Factor 4. Loadings on factor 4 suggest that schools that have greater proportions of underrepresented minorities as undergraduate medical students also have more women medical students and more minority faculty. The mean Medical College Admission Test (MCAT) scores of first year students at these schools also tend to be lower than those of other schools.

Factor 5. Three measures of success in obtaining single-investigator research funding from the National Institutes of Health (NIH) appear, again, to be independent of measures of a medical school's research emphasis and scope. The weak relationship of these variables to the proportion of women among medical schools, observed in the 1976 study, was not substantially different in 1977.

Factor 6. Data describing comparative characteristics of the curriculum were observed in 1975 and 1976 to be in short supply in the AAMC database. Selected variables that did discriminate among schools, those describing the use of elective courses, again were found to be independent of other principal components of institutional variation.

Factor 7. Four variables appear to indicate variability in the stage of development of medical schools. Younger medical schools tend to have fewer

students and anticipate expansion. They also are the schools that experienced the greatest recent increase in research funding from federal sources.

Factor 8 and Factor 9. The last two principal components relate to the research emphasis of medical schools. They are not easily distinguishable. The first characterizes schools that have large budgets, lesser reliance on volunteer faculty (who are usually involved in clinical activities), more NIH grants, and proportionately more medical students with excellent pre-medical grades. These variables may characterize a relatively large research enterprise. The second group of variables consists of only two with highest loadings on the ninth factor, but it also has several other variables with sizeable secondary loadings. The main variable is the number of basic science graduate students (usually Ph.D. candidates) per undergraduate medical student. Other related variables include the percentage of total institutional expenditures for sponsored research and the number of NIH grants approved. The common characteristic may be a research training mission of the school.

Cluster Analysis Results

On the basis of the results of the factor analysis described above, factor scores were produced for all schools which had data for at least 80 percent of the variables (27 variables). The decision rule for dropping schools was the same as that used in the earlier studies, and was employed to maintain consistency in the methods used. Two schools which were represented in the 1976 study, University of Washington and Rush Medical School, were not included. Four schools previously excluded for lack of data but included in this replication were Baylor University, North Dakota, University of Hawaii, and Eastern Virginia Medical College. As a result of the changes described above, 111 schools were included in the cluster analysis replication; in 1976, 109 schools were clustered.

Several combinations of factor scores were tried as input to the Ward hierarchical cluster analysis. All results were assessed for similarity to those found in the 1976 study. As in the previous study, the factor scores for the "minorities" and "elective courses" added no inherent meaning to the results and were again

dropped from the analysis. The factor which was labelled "alumni specialization" was also found to cause an apparent distortion in the cluster solution and was also eliminated from the final analysis. A final set of six factor scores was found to produce a hierarchical clustering which resembled the pattern derived in the earlier study. Not surprisingly the six factor scores selected as the basis for the final cluster analysis tended to replicate those of the 1976 study. The factor scores used in this study, then, were (1) graduate medical eduction, (2) control, (3) research funding success, (4) development, (5) research enterprise, and (6) research training.

The results of the hierarchical cluster analysis (which can be found in Appendix C) were an intermediate step in the anlysis. They served two purposes. In addition to aiding in the selection of the variables (factors) which would determine the composition of the eight clusters, the hierarchical cluster anlysis provided an indication of schools which could be used as initial seed points for the formation of clusters by the second step.

Using the information provided by the hierarchical cluster analysis, the 111 medical schools were grouped using non-hierarchical clustering based on the same six factor scores used in the first step. The results of the non-hierarchical cluster analysis are presented in Figure 1. The schools in each cluster are listed in the left column of Figure 1. The schools are listed in ascending order of their distance from the center of the cluster, and the number of parentheses is the Euclidian distance from a school to the center of the cluster. Schools which are closer to the center of the cluster are more typical of the profile of the cluster on the six factor scores. The cluster profile is depicted in the six boxes to the right of each list of schools. The rectangle within each box represents the mean of the schools in the cluster on the particular factor score. Since the factor scores are standardized with a mean of 0 and a standard deviation of 1.0 for all schools, a rectangle extending above the horizontal line indicates that a cluster has a mean on the factor score above the mean for all schools. Conversely, a rectangle extending below the horizontal line indicates that the schools in that cluster have a mean value which is below the mean of all schools. The

distance that the rectangle extends above or below the horizontal line is indicative of the degree to which the schools in the cluster are above or below the mean of all schools on a particular factor score, hence the degree to which the cluster is characterized by that factor.

In order to understand more fully the differences among clusters it may be beneficial to review the meaning of the six factors which were used in the analysis.

- Schools which have high scores in the graduate medical education factor would be those schools which had high ratios of housestaff (residents and interns) to undergraduate medical students, relatively high faculty salaries, high percentages of faculties who have M.D. degrees, and low ratios of undergraduate medical students to full-time faculty.
- Schools with high scores on the control factor are those which resemble the majority of private schools in that they have higher tuitions, relatively more out-of-state and foreign first-year medical students, and more applicants per matriculant; these schools also tend to receive a higher percentage of their revenue from the federal government.
- Schools which have high scores on the research funding success factor are those which have high approval rates on applications for research grants from NIH, are awarded higher percentages of the money requested through these applications, and have better (lower) standardized priortiy scores on their applications.
- Newer, smaller schools which have increased their degree of federal support over the past five years and plan to increase their enrollment of undergraduate medical students over the next five years would receive high scores on the development factor.
- Schools which have high scores on the research enterprise factor could be characterized as having large numbers of R01 grant applications approved, large amounts of revenue, small percentages of revenue devoted to administration and general expenses, fewer volunteer faculty per full-time faculty, and high percentages of first-year medical students having high pre-medical grande point averages.

FIGURE 1 CLUSTER MEMBERSHIPS AND CLUSTER PROFILES ON SIX FACTOR SCORES

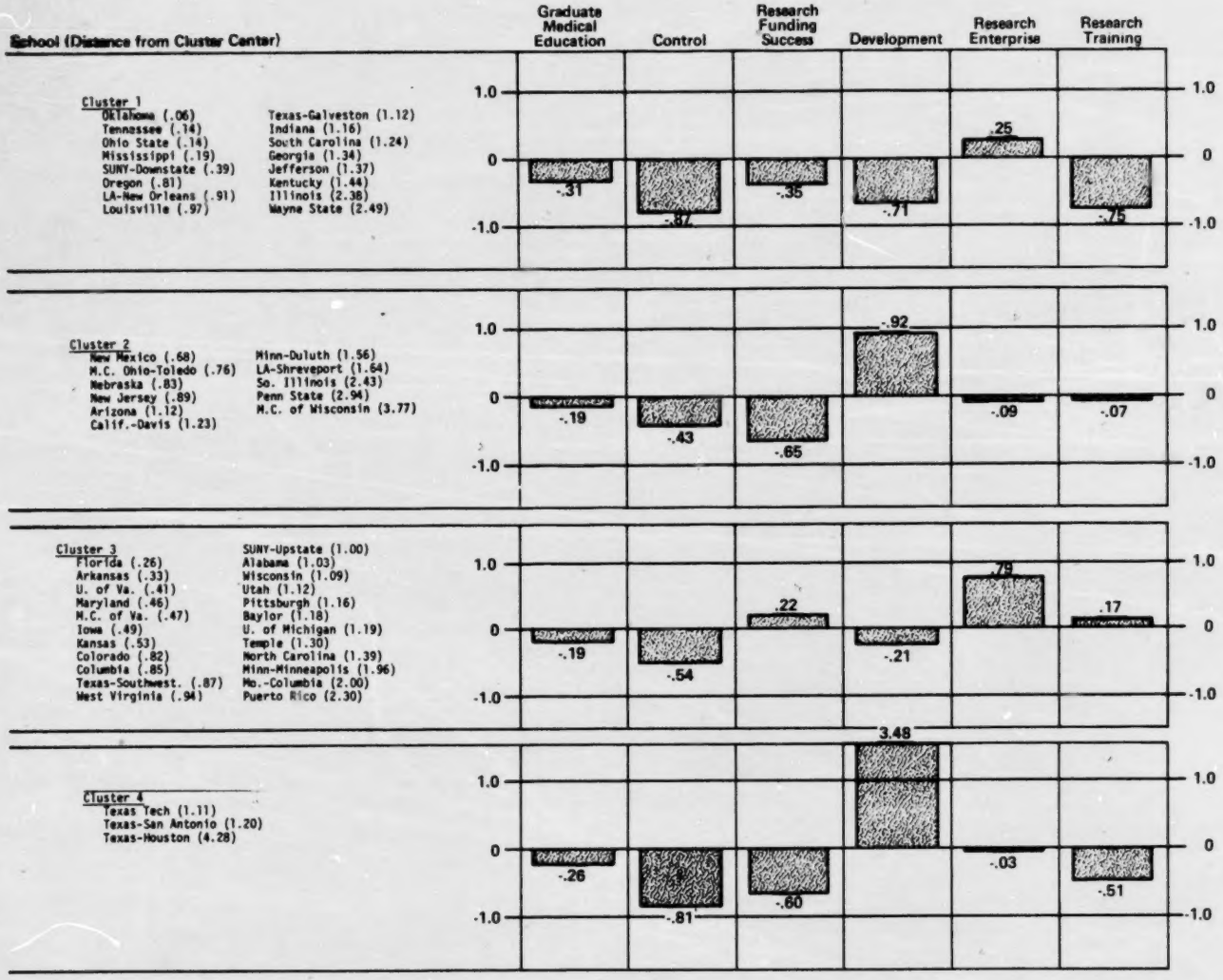
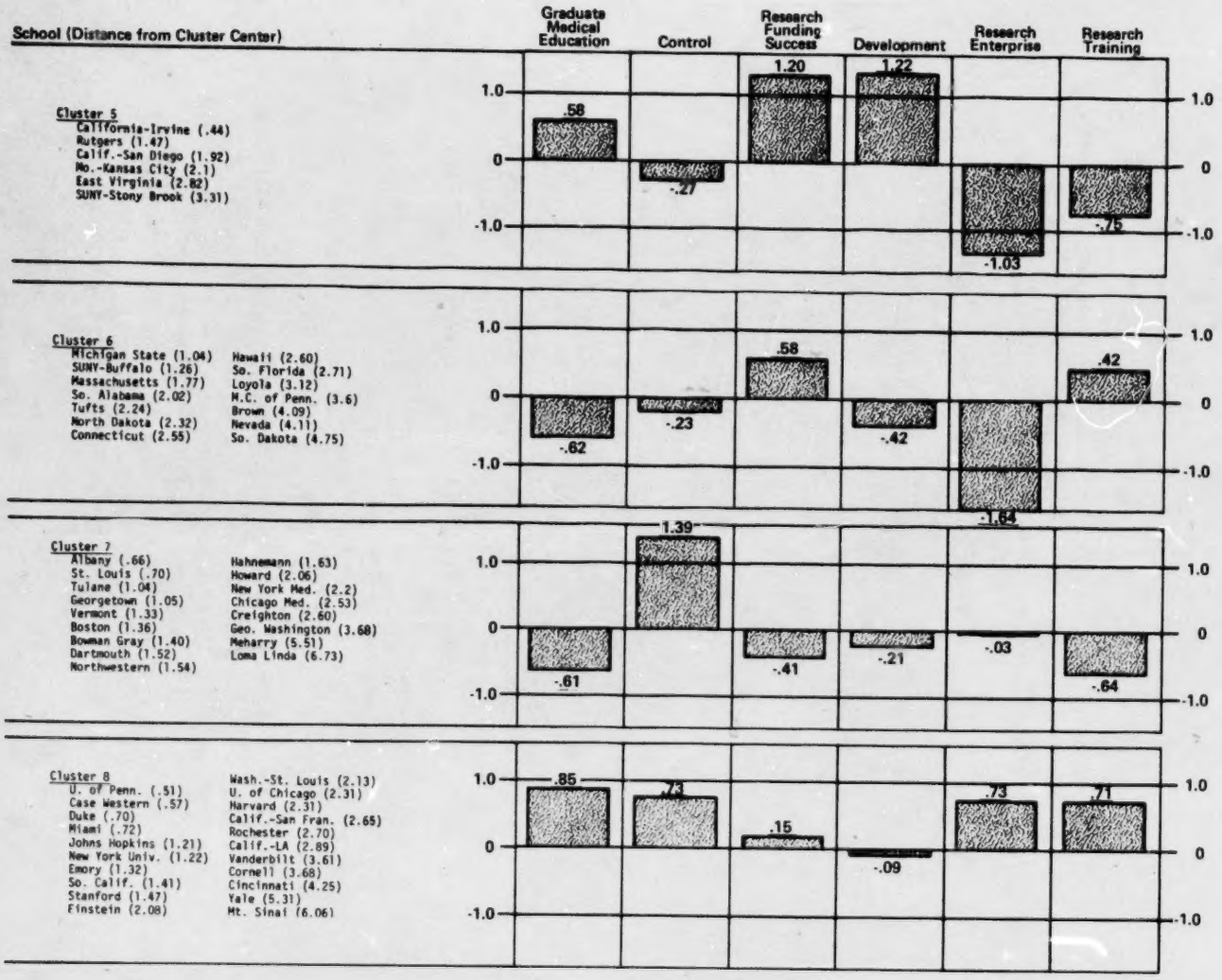


FIGURE 1 Continued



- **Schools with high scores on the research traning factor would have high ratios of basic science graduate student to undergraduate medical students and devote high percentages of their expenditures to sponsored research.**

To further aid the understanding of the composition of the eight clusters, the means for the eight clusters on the 25 variables which had the highest loadings on the six factors used in the cluster analysis are presented in Table 5. The variables in Table 5 are arranged in the order they appeared in the factor pattern matrix (Table 4), and the horizontal lines separate variables which had highest loadings on different factors. The first five variables had highest loadings on the graduate medical education factor, the next six on control, three on research funding success, four on development, five on research enterprise, and the final two on research training.

The first cluster depicted in Figure 1 is composed of one private and 15 public schools which tend to be in the middle of the distribution of schools on just about every facet examined in these analyses. The schools in Cluster 1 are slightly below average in graduate medical education emphasis and research funding success, slightly above average on research emphasis, and moderately below average on development and research training. When the cluster means in Table 5 are considered, it is apparent that these schools have a high ratio of undergraduate medical students to full time faculty and a low ratio of house-staff to undergraduate medical students. They also have the largest average number of undergraduate medical students of any of the clusters, and the lowest ratio of applicants to matriculants. These figures would indicate that the schools in Cluster 1 are predominantly oriented toward undergraduate medical education, and have relatively little emphasis on graduate medical education or research.

The schools in Cluster 2 are also primarily public schools (with one exception, Penn State). These schools, however, are somewhat newer and smaller than the schools in Cluster 1, and they are in a relatively early stage of development. These schools have relatively little research funding success compared to schools in other clusters, and are about average in research enterprise, research training, and graduate medical education. When individual variables are con-

sidered, the schools in Cluster 2 were found to have the lowest mean values of the schools in any cluster on the three variables associated with research funding success: the percentage of research funds applied for that were actually awarded, the approval rate of R01 grant applications, and mean standardized priority scores for R01 applications (where high scores indicate low priority). However, these schools have experienced a definite increase in research support from federal agencies between 1967 and 1974.

Cluster 3 is the largest cluster in this set, containing 23 medical schools, 20 of which are state supported. The domination of public schools is reflected in the moderately low score for the schools in this cluster on the control dimension. On four of the other five factor scores, this group of schools is close to the average of all medical schools. The only exception is the research enterprise factor, on which these schools are above average. This pattern is also found in the means of the variables used in the factor anlysis. While the schools in Cluster 3 have moderate values on most variables, they are among the highest in adjusted total revenue, number of R01 grant applications approved, and percentage of first year medical students with high pre-medical grade point averages. They also have the lowest mean values of any of the clusters on the percentage of expenditures devoted to administration and general expenses and the ratio of volunteer to full-time faculty.

In contrast with the cluster described above, Cluster 4 is the smallest cluster in the set, containing only three schools which are all public and all located in the state of Texas. The most remarkable characteristic of these schools is their growth, both in terms of increase of federal research support and in proposed increase in the size of incoming medical school classes. It is interesting that although these schools have enjoyed a remarkable growth in federal funding over the period from 1967 to 1974, they are below average, as a group, on research funding success. This contrast leads to the speculation that these schools probably started with very little federal research support in 1967-69, so that a small amount of growth was given added impact.

TABLE 5

MEAN VALUES FOR 8 CLUSTERS ON SELECTED VARIABLES CONTRIBUTING TO FACTOR SCORES USED AS THE BASIS FOR CLUSTER ANALYSIS

Variable	Label		CLUSTER 1	2	3	4	5	6	7	8
STC043	RAT: HOUSESTAFF TO UNDERGRAD MD STUD	Mean	.51	.66	.72	.70	.98	.58	.53	1.40
		(N)[1]	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
FAR053	FSS: AV SAL SFT ASSOC PROF BASIC SCI	Mean	24862.13	25749.70	25742.41	28729.67	30578.00	24548.00	24146.19	28047.68
		(N)	(15)	(10)	(22)	(3)	(4)	(12)	(16)	(19)
INC012	% REV FROM GIFTS & NON-GOVT SPONS PGMS	Mean	5.84	4.67	4.34	2.73	4.14	6.40	6.34	10.50
		(N)	(14)	(11)	(19)	(3)	(6)	(13)	(17)	(21)
INC058	RAT: MD STUDENTS TO FT FAC	Mean	2.26	1.54	1.42	1.21	1.62	2.66	2.27	.89
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
FAC001	FRS-% FT SAL FAC WITH MD OR MD-PHD	Mean	59.34	55.59	60.33	60.22	69.45	49.29	65.32	68.68
		(N)	(15)	(11)	(23)	(3)	(6)	(13)	(17)	(21)
STR432	1976-77 RESIDENT TUITION	Mean	1181.63	1692.70	1679.05	366.67	1785.67	2267.17	5365.43	3792.59
		(N)	(16)	(10)	(21)	(3)	(6)	(12)	(14)	(17)
STC029	% IN-STATE 1ST-YR MD STUD	Mean	94.33	92.21	89.33	97.22	89.64	80.98	39.43	54.03
		(N)	(15)	(11)	(21)	(3)	(6)	(14)	(17)	(19)
INR002	CONTROL: 0 = PUBLIC, 1 = PRIVATE	Mean	.06	.18	.17	.00	.17	.29	.94	.86
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
STC084	RAT: APPLICANTS PER 1ST-YR MD STUDENT	Mean	9.91	19.11	15.13	25.16	33.35	23.36	42.29	40.15
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
INC007	% REV FROM FED SOURCES & RCOV IND COSTS	Mean	27.63	27.53	35.46	15.17	33.36	29.26	39.25	46.28
		(N)	(16)	(11)	(19)	(3)	(6)	(12)	(14)	(21)
STC008	% NON-US-CANADIAN 1ST-YR MD STUDENT	Mean	.26	.06	.78	.53	.60	1.44	2.00	2.54
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)

TABLE 5 (Cont'd.)

Variable	Label		CLUSTER 1	2	3	4	5	6	7	8
INR141	NIH RO1 GRANTS: $ AWARD AS % OF APP SBMT	Mean	19.49	15.05	25.25	19.57	31.97	25.14	19.69	26.29
		(N)	(16)	(10)	(22)	(3)	(6)	(14)	(17)	(21)
INR137	NIH RO1 GRANTS: APPROVAL RATE OF APPLS	Mean	58.58	53.64	75.27	59.00	66.35	78.76	58.84	75.40
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
INR142	NIH RO1 GRANTS: MEAN STD PRIORITY SCORE	Mean	.22	.34	.04	-.03	-.66	.05	.33	-.09
		(N)	(16)	(9)	(23)	(2)	(6)	(14)	(17)	(21)
INC003	DRG FED SPON RES CON$ CHG 67-9 TO 72-4	Mean	-7.83	173.85	16.29	577.83	177.60	27.06	16.49	21.14
		(N)	(16)	(8)	(23)	(2)	(3)	(10)	(17)	(21)
STC114	PROJTD ANNL % 1ST-YR ENROLL CHG 1975-80	Mean	.34	3.47	.27	17.32	9.79	.52	.34	-.09
		(N)	(16)	(8)	(22)	(3)	(5)	(14)	(17)	(21)
INC048	LOG AGE OF MEDICAL SCHOOL	Mean	1.95	1.19	1.95	.92	.98	1.43	2.04	1.87
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
STR009	# UNDERGRADUATE MEDICAL STUDENTS	Mean	750.81	317.00	578.87	264.33	232.17	314.00	526.35	515.86
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
INC004	ADJUSTED TOTAL REVENUE THOUSAND $	Mean	30705.01	20071.23	40408.98	24238.59	17359.64	10435.09	19627.89	53573.78
		(N)	(16)	(11)	(21)	(3)	(6)	(14)	(17)	(21)
INC026	% EXPD FOR ADMIN & GENL EXPENSE	Mean	11.54	9.82	7.58	15.42	12.05	15.42	10.59	8.13
		(N)	(16)	(11)	(21)	(3)	(6)	(14)	(17)	(21)
FAC019	RAT: VOL FAC TO FT FAC	Mean	2.16	2.48	1.39	3.43	5.45	4.21	1.87	1.54
		(N)	(16)	(11)	(22)	(3)	(6)	(14)	(17)	(21)
INR136	NIH RO1 GRANTS: # APPLS APPROVED	Mean	19.44	13.45	33.26	11.33	14.67	12.43	13.59	45.67
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
STC013	% 1ST-YR MD STUD: PRE-MED GPA 3.6-4.0	Mean	46.83	41.81	49.34	15.68	32.75	31.16	31.96	59.29
		(N)	(15)	(11)	(22)	(3)	(6)	(14)	(16)	(19)
STC045	RAT: BMS GRAD STUD TO UNDERGRAD MD STUD	Mean	.19	.18	.32	.28	.18	.25	.15	.44
		(N)	(16)	(11)	(23)	(3)	(6)	(14)	(17)	(21)
INC017	% TOTAL EXPD FOR SPON RESEARCH	Mean	14.71	13.04	21.69	10.66	17.97	16.71	20.31	35.41
		(N)	(16)	(11)	(21)	(3)	(6)	(14)	(17)	(21)

1. Numbers in parenthesis are number of schools having data for that variable.

Cluster 5 is another small cluster, composed primarily of public schools. Although the schools in this cluster are also developing, they are also notable for a high degree of research funding success and a higher than average involvement in graduate medical education. This is in contrast to their below average means on research enterprise and research training. This group of schools has the highest mean percentage of funds awarded in R01 grant applications and the highest mean standardized priority scores of any of the clusters. They also had a high average ratio of housestaff to undergraduate medical students, the highest average salary faculty and the highest percentage of faculty with M.D. degrees of any of the clusters.

Cluster 6 is a fairly large group of schools also dominated by public schools. As a group these schools are somewhat below average in graduate medical education and are not in early development stages. However, it is of interest to note that although these schools have moderately high mean values on research funding success and research training, they have the lowest mean value of all the clusters on research enterprise. In terms of the actual variables, these 14 schools have the lowest adjusted total revenues of any of the groups, the lowest average faculty salaries, and the lowest percentage of faculty with M.D. degrees of any of the clusters. These values may be due in part to the fact that a few of the schools in this cluster are relatively new, and that three of the schools (North Dakota, South Dakota, and Nevada) are in some stage of the transition from two-year basic science schools to M.D. degree-granting institutions.

The final two clusters are both composed largely of private medical schools, but have very different characteristics. The schools in Cluster 7 are below the average of all schools as a group on every factor other than control. They are slightly below average on development and research enterprise, moderately below average on research funding success, and substantially below average on graduate medical education emphasis and research training. It appears that these schools concentrate on their undergraduate medical education programs, because they have very low values on the ratios of both housestaff and basic science graduate students to undergraduate medical students. These schools are, on the average, the oldest, the most

expensive to attend, have the highest ratio of applicants to matriculants, and have the lowest percentage of in-state first year medical students.

The final group of schools, Cluster 8, has the highest profile of values on the variables used in this analysis of any of the clusters. In addition to having moderately large undergraduate medical student enrollments, these schools have graduate medical education, research enterprise and research training components that are well above the average for all schools. As a group these schools have the largest adjusted total revenue, the largest number and highest percentage of R01 grant applications approved, the highest percentage of first-year medical students with high pre-medical grade point averages, and the highest ratios of both housestaff and basic science graduate students to undergraduate medical students of any of the clusters. In addition, these schools receive the highest percentage of the revenue from the federal government and from gifts and non-government sponsored programs, and enroll the highest percentage of foreign first year medical students. They also have the lowest ratio of undergraduate medical students to full time faculty and the lowest projected growth rate over the next five years of any of the clusters.

Comparison with 1976 Results

While the clusters found in the 1977 replication do form meaningful groupings, one of the objectives of this study was to determine whether the repetition of analyses on newer data would yield the same clusters found in 1976. Figure 2 presents a comparison of the schools in each cluster in the 1976 study and the 1977 replication. Had the clusters in the 1976 study been exactly reproduced in the 1977 replication, all of the schools would have fallen in the cells along the diagonal running from the top left to the bottom right of Figure 2. It is obvious that the 1977 replication did not reproduce the 1976 study clusters.

In spite of the imperfect replication, there are a number of similarities in the structure of the clusters that were present in both years. In both years, for example, public and private schools tended to be placed into separate clusters. In addition, newer, developing

FIGURE 2

COMPARISON OF CLUSTER MEMBERSHIPS BETWEEN 8 CLUSTERS IN 1976 STUDY AND 1977 REPLICATION

1977 REPLICATION CLUSTERS

1976 STUDY CLUSTERS	1	2	3	4	5	6	7	8
1	Georgia, Ken. LA-N.O., Tenn. Louisville Miss., S. Car. Ohio State Okla., Ore.	Nebraska	Arkansas Maryland M.C. of Virginia		Rutgers	South Dakota South Alabama		
2	Illinois Indiana Jefferson Suny Downstate Texas Galveston Wayne State		Temple			Suny Buffalo		
3		New Mexico	Alabama, Iowa Color., Kansas Fla., N. Car. Mo.-Columbia P.R., Utah, Wis. U. of Va., W.V.			Connecticut		
4		New Jersey	Columbia U. of Michigan Minn.-Minneapolis Pittsburgh Texas Southwest Suny Upstate				New York Med.	Calif., San Fran Calif., L.A. Einstein Harvard Miami New York University
5		So. Illinois			Missouri K.C.	Nevada	Chicago Medical	
6		Penn State Calif. Davis MC Ohio Toledo L.A.-Shreveport		Tex San Antonio Texas Houston Texas Tech	Calif, Irvine Cal. San Diego Suny Stony Brk	Massachusetts Michigan St. South Florida		
7						Loyola Stritch Tufts MC of Penn. Brown	Albany, BU, GWU Bowman-Gray Creighton, GU Dartmouth, N.W. Hahnemann Meharry Loma Linda, St.L Tulane, Vt., HU	
8		MC of Wisconsin						Chi, Cornell, Mt. Sinai Cincy, Stanford, Case W Duke, Emory U.Pa, St. L Johns Hopkins, Yale Roch. S. Ca., Vanderbilt

schools tended to be separated from older, established schools. Closer examination of Figure 2 reveals that Cluster 1 and Cluster 8 in the replication were formed entirely by schools in two clusters each in the 1976 study (1 and 2, and 4 and 8 respectively). Cluster 2 in the replication is a recombination of a number of developing schools from the 1976 study with a number of older, established schools from other clusters. Each of the four established schools added to this cluster was an outlier (only weakly associated) in the 1976 cluster of which it was a member. Cluster 3 in the replication is composed primarily of schools in Clusters 3 and 4 in the previous study. Clusters 4, 5, and 6 were predominantly a reorganization of developing schools, which could be due to real changes in the interrelationships among the schools on these variables. Finally, Cluster 7 is almost a reproduction of Cluster 7 from the 1976 study, with two schools added and four splitting off to join other clusters.

In summary, the results of these studies indicate that although the milieu in which medical schools exist is not constant, there are certain patterns among variables and among institutions which appear to remain stable from one year to the next. These relationships could form the basis of future studies.

Chapter IV

SUMMARY AND CONCLUSION

This report describes the replications of the principal components analysis and cluster analysis which were performed in 1976. The replications, wherever possible, used the same variables which had been used in 1976 but with data one year more recent. Identical analytic techniques were used in both years. The results of these exercises showed that there were many similarities among the patterns found across two years, but that the 1976 results were not exactly reproduced by the 1977 studies.

There seems to have been a pyramiding effect in the replication analyses. The principal components analysis obtained a pattern of relationships among the variables which was somehow different than the 1976 pattern. Therefore, the factor scores computed for each school, which are sensitive to changes in the relationships among all variables, represented somewhat different theoretical dimensions than in 1976, and, due to the sensitivity of cluster analysis, the derived clusters were modified substantially. A less severe, and possibly more fair, test of the cluster analytic results found in 1976 would have been to compute the replicated factor scores based on 1976 weights with 1977 data, and to use those factor scores as input to the cluster analysis. This would have allowed for use of an established set of weights, and the results of the analyses would not have been subject to the pyramid effect described above. However, such a scheme would make use of relationships among variables which may not hold from one year to the next.

There is one additional aspect of this work which merits consideration. It would be useful in further work of this nature to assure that all variables and indices selected for analysis on a repetitive basis could be collected and updated annually. As described in Chapter II, three of the variables used in these studies were not recurrent, and the effects of having variables which were not repetitious has been noted. Essentially non-recurrent data forces a choice of using exactly the same values in successive years or modifying in some way the variable selected. Both approaches were used in this replication, and neither was found to be entirely satisfactory.

In conclusion, it seems evident that the line of inquiry followed in this series of studies thus far has revealed a number of interesting patterns, both expected and surprising, among variables and institutions. It would be valuable, however, for future studies to be limited to recurrent variables. In this way, the results of this work may lead to stable and useful institutional "indicators" that could be used to accurately monitor the continual changes in institutions of medical education in the United States.

BIBLIOGRAPHY

Agro, G., Hilles, W., McShane, M., Rosenthal, J. Institutional Characteristics of U.S. Medical Schools, 1975-76. Washington, D.C.: AAMC, 1977.

Forgy, E. W. "Cluster Analysis of Multivariate Data: Efficiency Versus Interpretability of Classifications." Paper presented at the Biometric Society Meeting, Riverside, California, 1965.

Martin, B. C. Medical School Alumni. Chicago: American Medical Association, 1975.

McShane, M. G. An Empirical Classification of U.S. Medical Schools by Institutional Dimensions. Hyattsville, Maryland: Bureau of Health Manpower, DHEW Publication No. (HRA) 77-55, 1977.

Sherman, C. R. A Second Exploratory Analysis of the Relations Among Institutional Variables. Hyattsville, Maryland: Bureau of Health Manpower, DHEW Publication No. (HRA) 77-57, 1977.

Ward, J. H. "Hierarchical Grouping to Optimize an Objective Function," Journal of the American Statistical Association, 58:236-244, 1963.

APPENDIX A

Abbreviations Used in Variable Labels

Symbol	Definition
#	Number
%	Percent
AMA	American Medical Association
ANNL	Annual
APP	Applications
APPLS	Applications
ASSOC	Associate
AV	Average
AWARD	Awarded
BAS SCI	Basic Science
BMS	Basic Medical Science
CHG	Change
CON$	Constant 1964 Dollars (by NIH R&D Index)
DRG	Division of Research Grants (at NIH)
ENROLL	Enrollment
ETHNIC	Afro, Chicano, American Indian, and Mainland Puerto Rican
EXPD	Expenditures
FAC	Faculty
FED	Federal Government
FR	From
FRS	Faculty Roster System
FSS	Faculty Salary Survey
FT	Full-Time
GENL	General
IND COSTS	Indirect Costs
MATRC	Matricultants, First Year Students
MCAT	Medical College Admissions Test
MD	Medical, Medical Degree
NON-GOVT	Not Local, State, or Federal Government
PGMS	Programs
PROF	Professor
PROJD	Projected
RAT	Ratio
RCOV	Recovery
REL	Related to Medicine
REV	Revenue
SAL	Salary, Salaried
SBMT	Submitted
SCR	Score
SFT	Strict-Full-Time
SPONS	Sponsored (restrict revenue)
STD	Standardized (within study group)

Symbol	Definition
STUD	Student
UNDERREP	Underrepresented (Afro, Chicano, Mainland Puerto Rican)
VOL	Volunteer
YR	Year

APPENDIX B

Interpretation of the Factor Pattern Matrix

An understanding of the interpretation of the numerical "loadings" that comprise the factor pattern matrix facilitates the assessment of the results of the factor analysis used for the exploratory purposes of these studies.

The numbers in a table of "factor loadings" are measures of strength of association between the variables and the derived "factors". Like correlation coefficients representing the relationship between pairs of simple variables, they range in value from +1.0 to -1.0. Values near zero represent "no relationship"; values near +1.0 or near -1.0 represent strong positive and strong negative relationships respectively. The first row shows how strongly the first variable is related to each factor. Because of the rotational criterion, any one variable is probably highly related to only one or two factors and weakly related, at best, to the other factors. For purposes of speculation it is assumed that variables related to the same factor are likely to be related to each other.

For ease of examination, the variables in the table are often ordered according to their highest factor loadings. The predominant loading (or loadings) for each variable are highlighted with a "box" (for high values) or an "asterisk" (for moderately high values). The grouping of variables means that they may be related to one another, that is, their values vary the same way across institutions. At any given school, high standardized values of one variable tend to be matched with high values of the other, low with low, if the relationship is positive, that is, if the signs on the loadings are the same (both "plus" or both "minus"). If the signs of two variables' loadings are different (one "plus" and one "minus") the relationship is probably negative, that is, high standardized values of one variable are matched with low values on the other. Because the factors are numerically independent of one another (due to the rotational procedure used), it is also likely that the variables in one group have low correlation with variables in another group. Exceptional variables are readily seen.

By way of additional guidance in the interpretation of the factor pattern matrices, two additional rules of thumb may be useful. First, factor loadings with value less than about .50 (in absolute value) should not be given as much attention as larger numerical loadings. Second, variable groupings that account for small percentages of overall variance (given at the bottom of each column) may be less accurate indicators of potential relationships than groupings accounting for greater percentages of variance.

Whereas the named "factors" may be conceptually or mathematically independent and most variables related only to one factor, some individual variables may be found to be related to more than one factor. This may be more easily understood through a simple analogy. If, instead of medical schools, rectangles were the unit of study, height, width, and area might be among the measured variables. As a result of analysis, height and width may be found in a common factor with area, but, since height and width are independent of each other, one or both may also be found in additional factors (variable groupings).

APPENDIX C

WARD HIERARCHICAL CLUSTER ANALYSIS - 1977 REPLICATION